

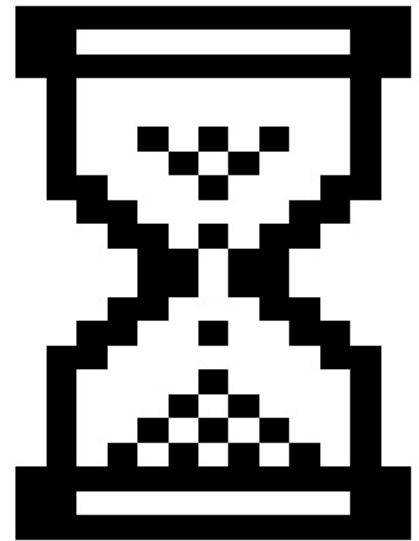
Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State

Avar Pentel

Tallinn University, Institute of Informatics,
Tallinn, Estonia, pentel@tlu.ee

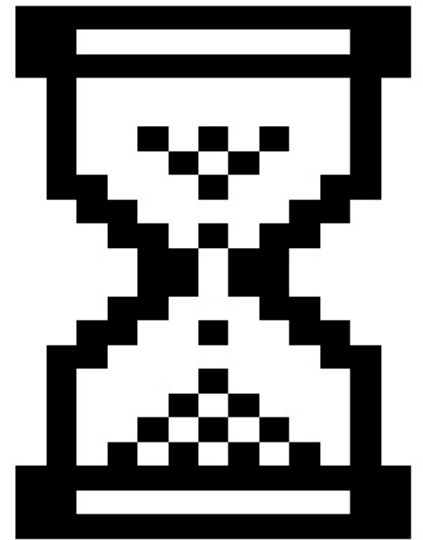
Intro

- Advice from a 8 year old boy: when your computer becomes unresponsive...



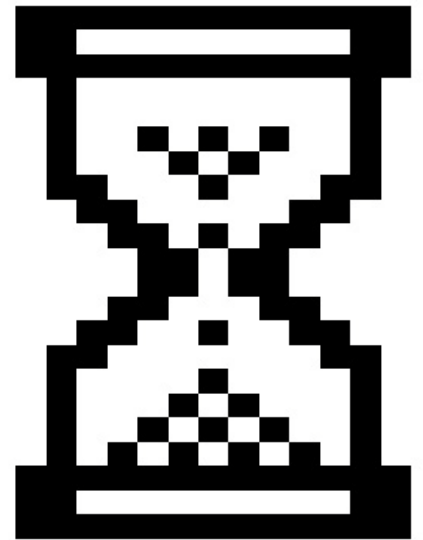
Intro

- Advice from a 8 year old boy: when your computer becamnes unresponsive...
- You have to move your mouse like this, then it will go faster.



Intro

- Advice from a 8 year old boy: when your computer becamnes unresponsive...
- You have to move your mouse like this, then it will go faster.



Emotions and body movements

are related

- Gesticulation
- Posture

Goal

To predict user's emotional state by analyzing mouse movement logs

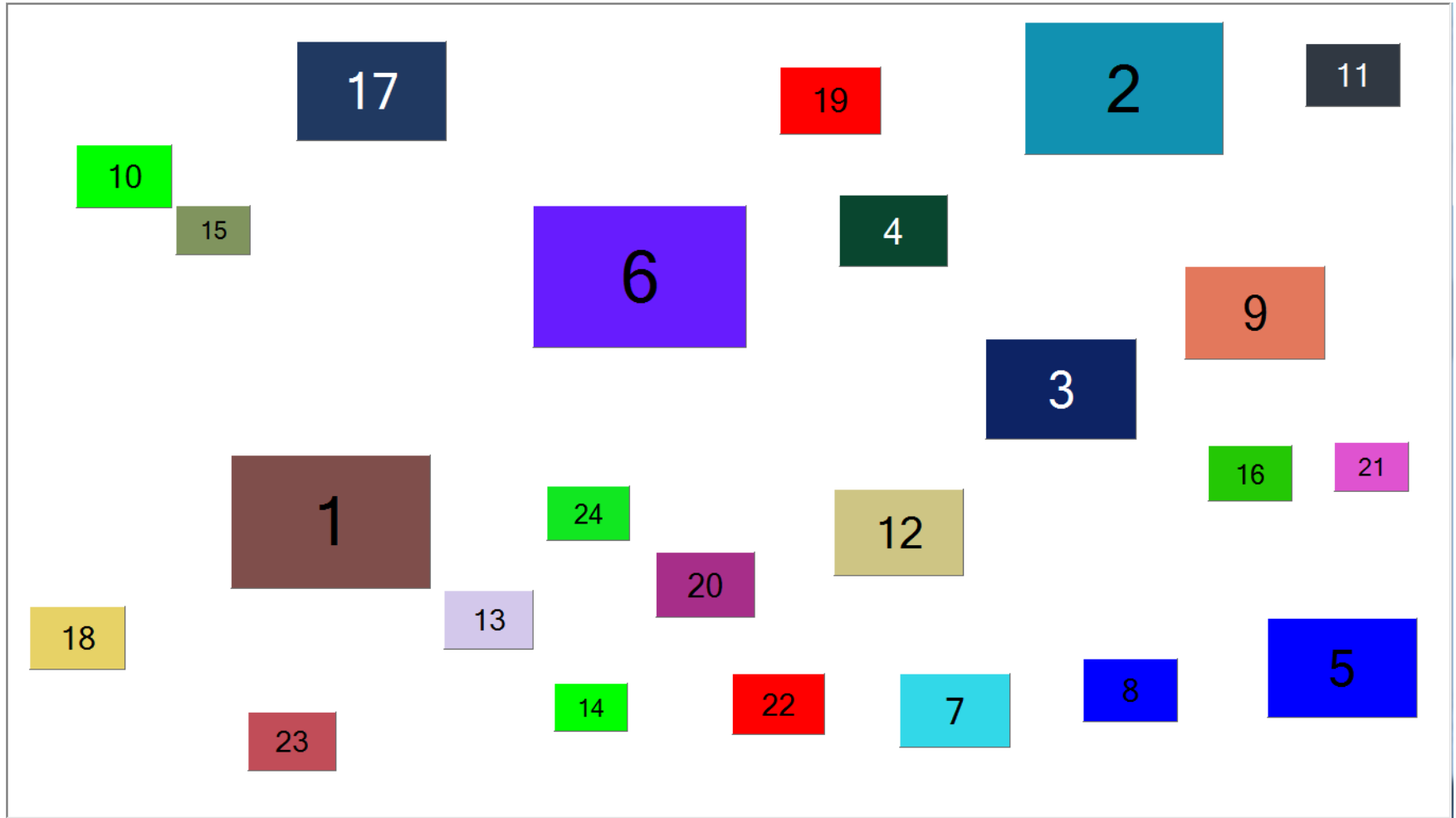
Outline of the Presentation

- Related work
- Experimental setup
 - Data collection procedure
 - Associating tasks to emotional states
 - Features
 - Machine learning
- Results
- Conclusion

Related work

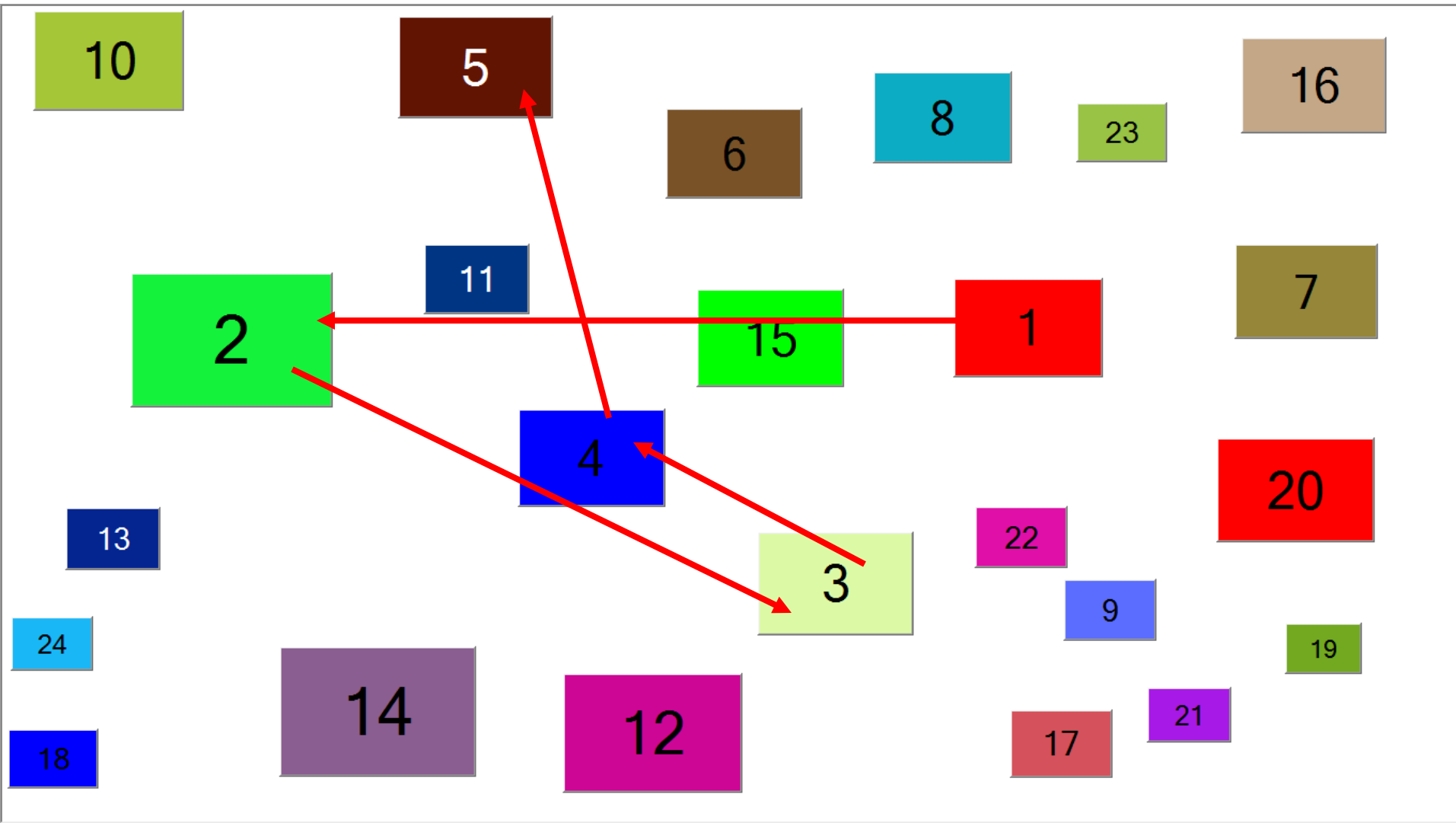
- Special equipment
- Small samples
- Specific tasks, no general link between emotion and mouse movement is studied

Data Collection Procedure



Idea from Christmas Calendar

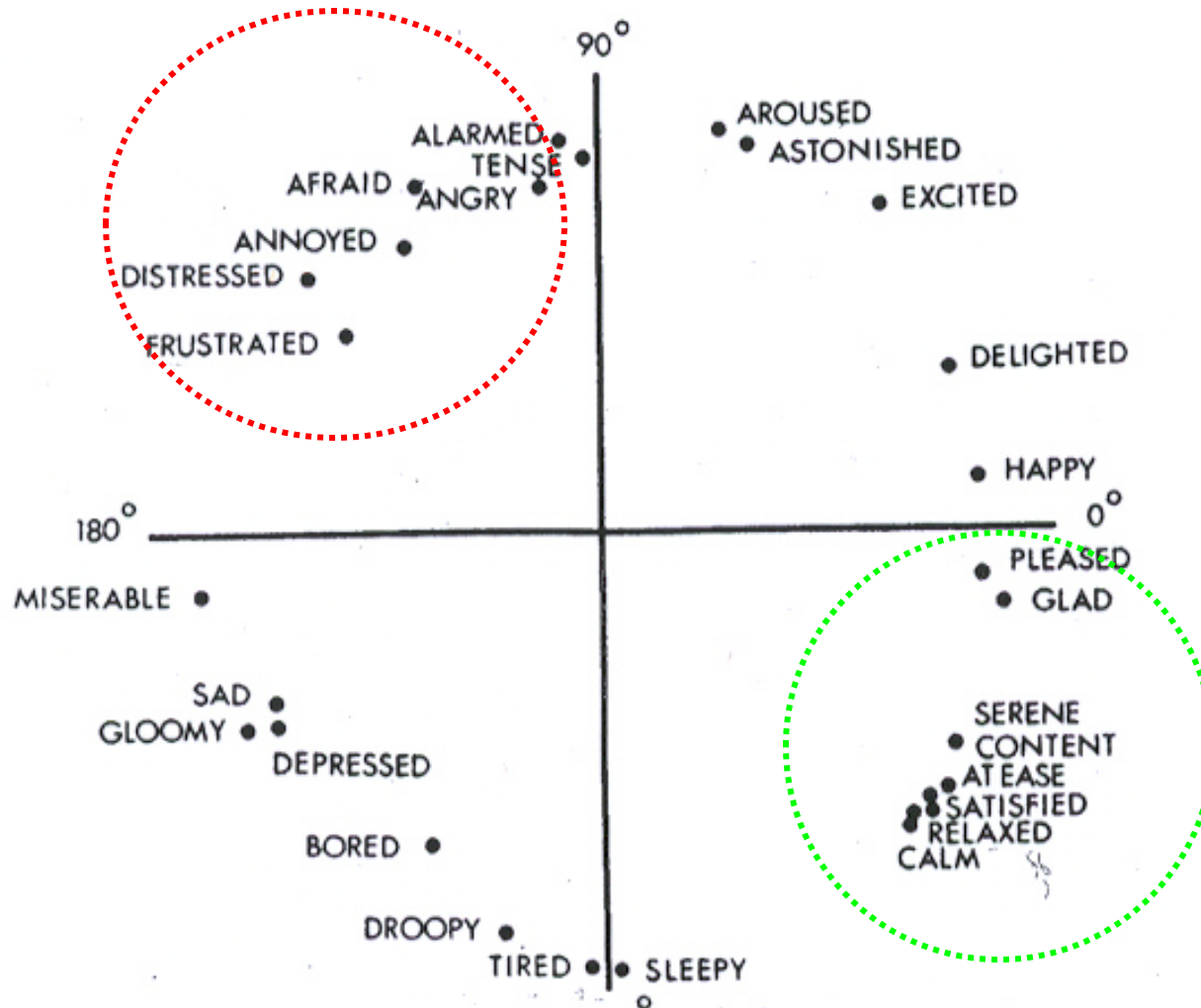




Data collection

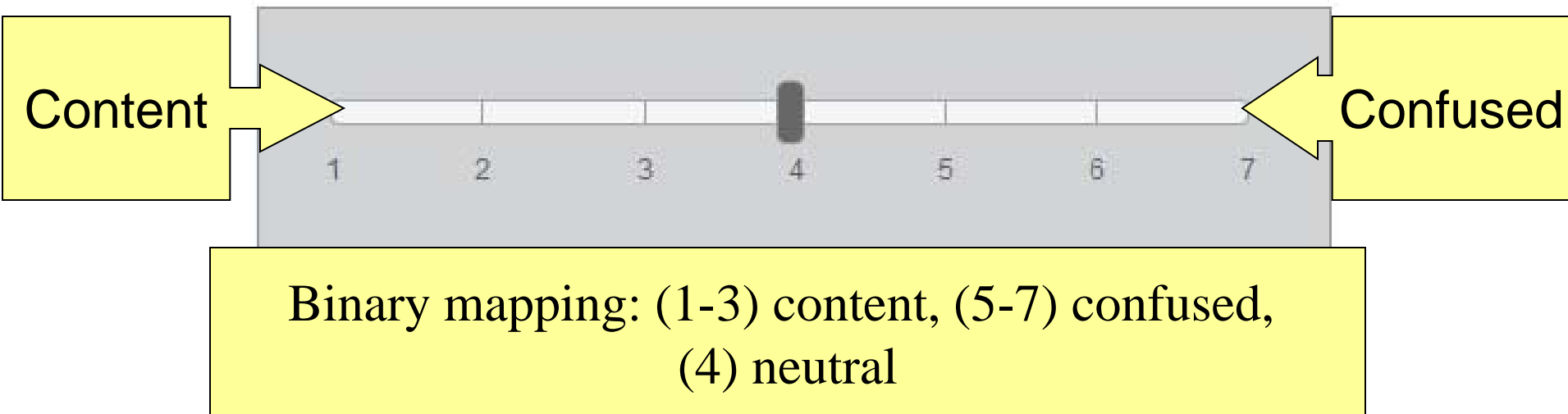
- The game has collected data about:
 - Each mouse click.
 - Time when button was clicked.
 - All mouse movements with timestamps
- 516 game sessions played by 262 individual users
- Each game session consisted of 24 searching tasks (to find the next number)
- 12384 comparable records (standardized session-wise), each of them presenting mouse movement logs between two button clicks

Self-Reports on Russel's Model



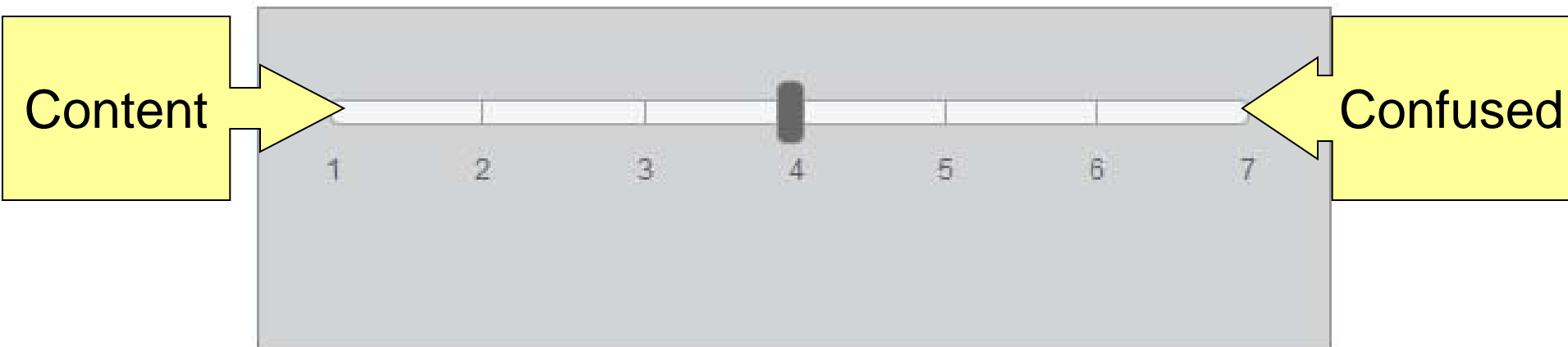
Self-Reports on Likert Scale

- Interviews with selected participants (N=44)
- Right after game session
- Still image of the game session was shown



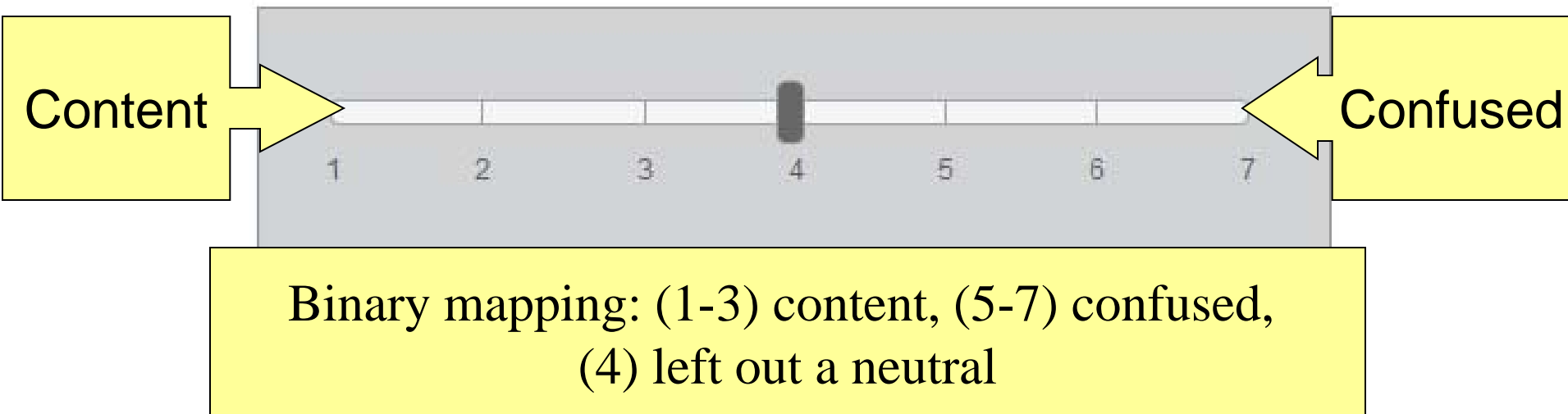
Self-Reports on Likert Scale

- Emotion data about $44 \times 24 = 1056$ tasks
- All target fingering times standardized session-wise
- Pearson correlation between self reports and standardized finding time was found ($r = 0.86$)

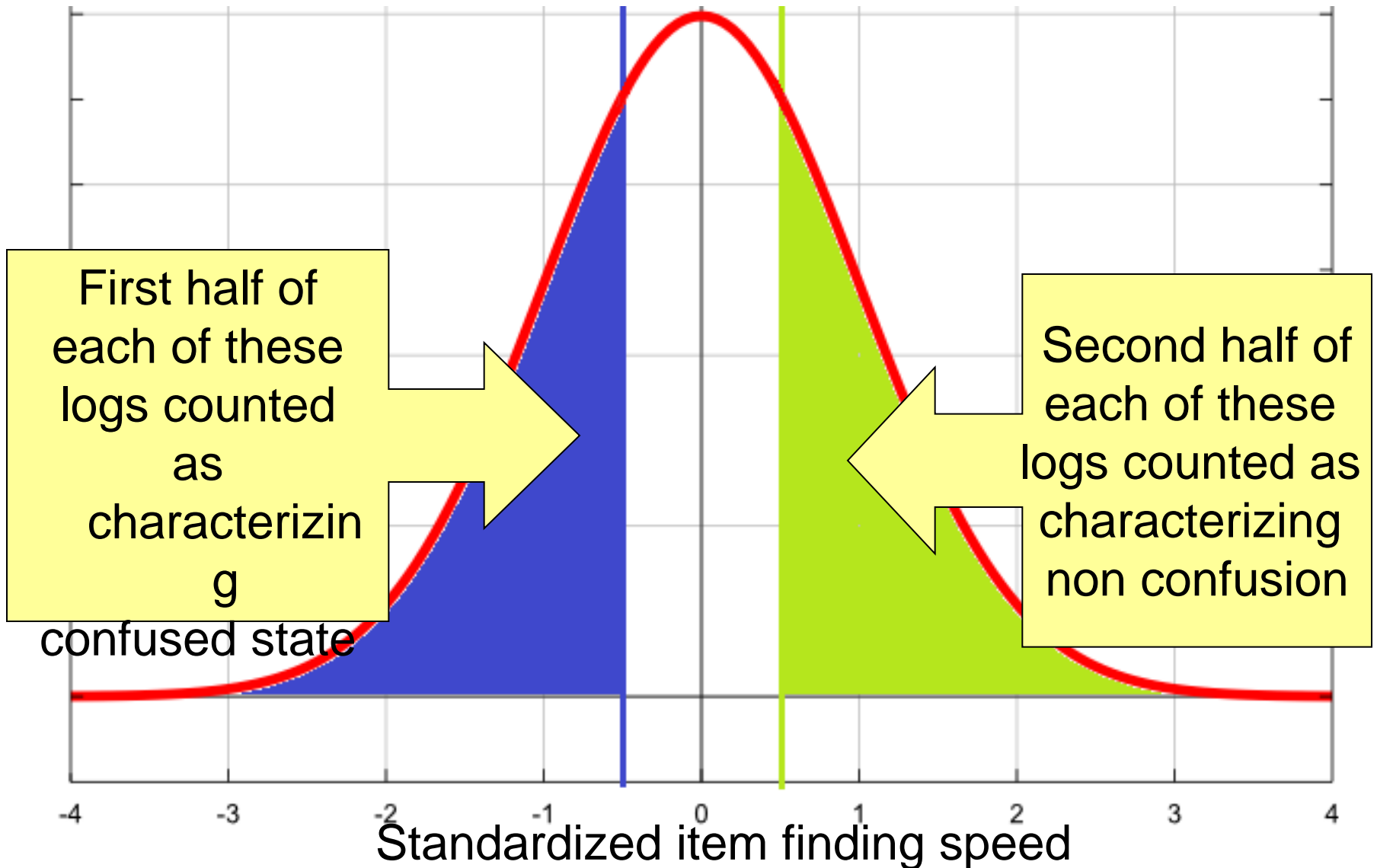


Self-Reports on Likert Scale

- Tasks reported as confused had finding speed 0.5 standard deviation below mean
- Tasks reported as content had finding speed 0.5 above mean



Separation of Classes



Final Dataset

- Out of these two subsets we excluded repeated sessions by the same users, extreme results, and game sessions that contained clicking errors.
- Out of the remaining data we created balanced training dataset of 2282 records

Features

Type	Feature	Explanation
Distance*	Precision	Shortest distance between two button clicks and actual mouse path length ratio.
Speed**	Speed	Actual mouse path length between two button clicks divided by task completion time.
	AdjSpeed	Actual mouse path length between two button clicks divided by shortest path, and then divided by task completion time.
Direction	DirectionX	Number of mouse movements in particular direction. We divided movement directions to 8 distinctive segments as north, northeast, east, etc. We counted all movements in particular direction segment, and divided to all movements.
Direction changes	TurnA	Mouse movements' path was recorded as consecutive straight lines of 10px length. We measured each angle between two consecutive movements and extracted 18 features representing turns from 0 to 180 degrees by 10-degree step. Counted results were normalized by whole number of movements.
	Turn10,	
	Turn20,...	
	Turn 180	
	TurnA+	All turns greater than angle A (A counted by 45-degree step).

* Excluded in training feature set of the models titled as “target unknown”:

** Excluded in all training feature sets.

Features

Machine learning

- Logistic Regression
- Support Vector Machine
- Random Forest
- C4.5
 - Motivation based on literature.
 - Java implementations of listed algorithms that are available in freeware data analysis package Weka
 - 10-fold cross validation

Results

Model	Target known			Target unknown		
	<i>Accuracy</i>	<i>F-score</i>	<i>ROC</i>	<i>Accuracy</i>	<i>F-score</i>	<i>ROC</i>
SVM (standardized)	94.61%	0.946	0.946	82.38%	0.824	0.825
Logistic Regression	93.49%	0.935	0.978	82.72%	0.827	0.889
Random Forest	92.07%	0.921	0.971	84.47%	0.845	0.825
C4.5	91.96%	0.919	0.937	83.59%	0.835	0.836

Conclusion

- Mouse movements reveal users emotional states
- But is the confusion and frustration in current study comparable with confusion and frustration caused by solving mathematical equation or some other cognitively more demanding task?

Conclusion

Different kind
of Confusions

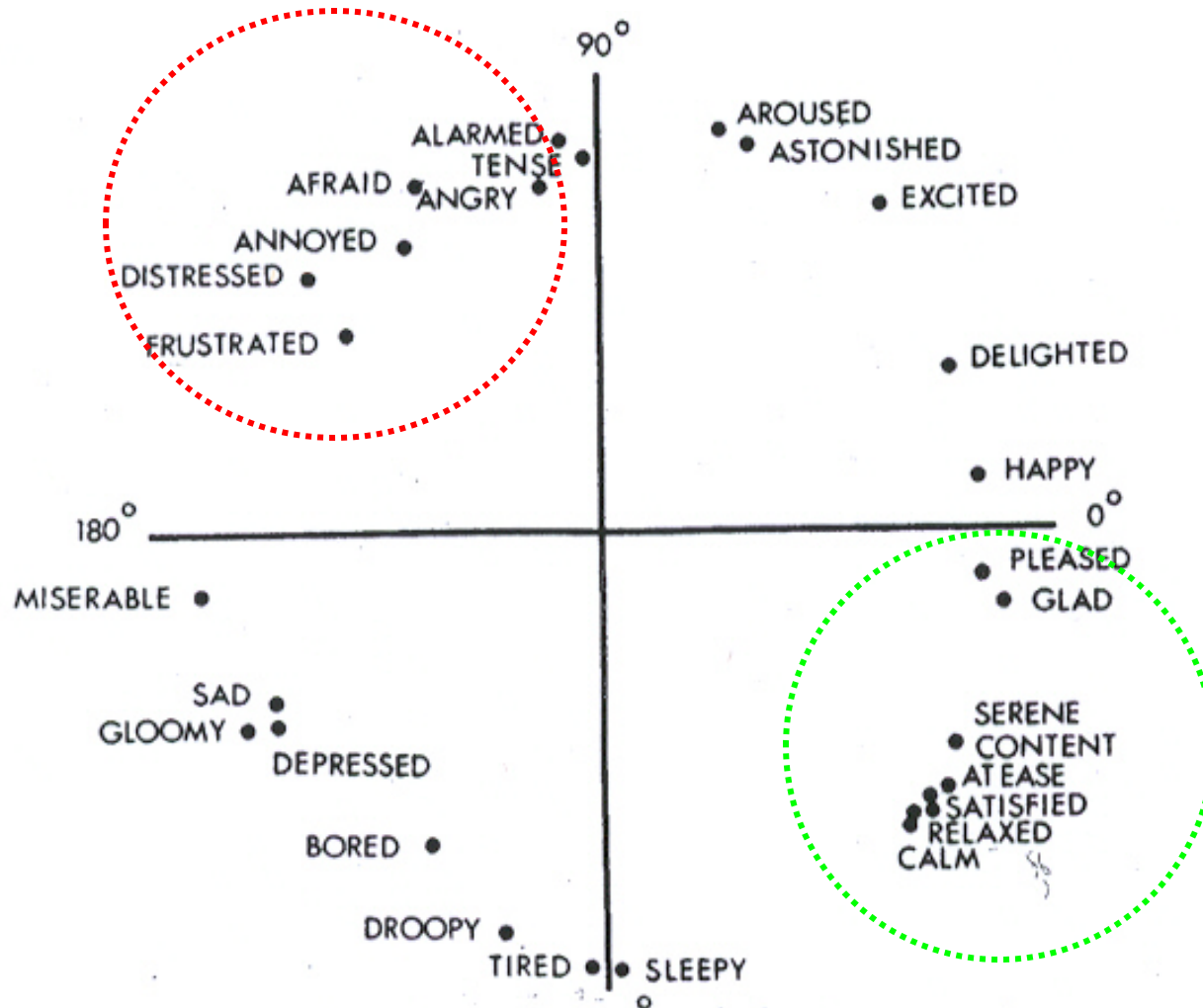


Conclusion

Different kind
of Confusions



Conclusion: Russel's model



Conclusion: Limitations

- Reliability of retrospective self-reports
- Too many items (24) to remember

Thank You!

Q&A